Joinpoint Trend Analysis of Cancer Incidence and Mortality using Alberta Data

Zhichang JIANG ● Zhenguo QIU ● Juanita HATCHER
And the Cancer Projection Analytical Network Working Team

Cancer Surveillance, Surveillance and Health Status Assessment, Alberta Health Services
1400 Sun Life Place, 10123 – 99 Street, Edmonton, Alberta, T5J 3H1
Tel: 780-6434384; Fax: 780-6434380
Email: zhichang.jiang@albertahealthservices.ca
Production of this document has been made possible through a financial contribution from Health Canada, through the Canadian Partnership Against Cancer.

The views expressed herein do not necessarily represent the views of the Canadian Partnership Against Cancer nor that of Health Canada.
Contents

I. Introduction ..........................................................................................................................4

II. Methods ............................................................................................................................7

   1. Regression Models ......................................................................................................... 7
   2. Joinpoint Regression Model ............................................................................................9
   3. Joinpoint Regression Software .......................................................................................12

III. Results and Conclusion ..................................................................................................16

   1. Data Preparation ..........................................................................................................16
   2. Parameter Setting in Joinpoint Software ........................................................................16
   3. Outputs ..........................................................................................................................18
   4. Summary Results ............................................................................................................22
   5. Results for Special Considerations .................................................................................25

IV. Discussion and Recommendations ....................................................................................34

V. Appendix ..........................................................................................................................38
I. Introduction

Health officials and policy-makers require information about the future cancer burden in order to assist in the planning and prioritization of prevention activities, allocation of health services, and evaluation of cancer control interventions or treatments. Although projections of cancer incidence and mortality are essential for the facilitation of these activities, there is currently no Canadian mechanism that will permit the systematic review of available projection methods or identification of efficient long-term projection methodology. A pan-Canadian Cancer Projection Network (C-Proj) was established as part of the Canadian Partnership Against Cancer Surveillance and Epidemiology Networks in order to address the need for comprehensive and consistent cancer projection information. The C-Proj Network is a collaborative and multi-disciplinary initiative that will evaluate and implement standardized methodologies to predict future cancer burden across the country. To estimate the future number of new cancer cases or deaths, an assessment of cancer incidence and mortality trends is necessary. The identification of past trends allows projected trends to be estimated by extrapolating time trends from observations. In this report, trend analysis techniques using Joinpoint software will be discussed and decision-making considerations elucidated.

Trend analysis is a technique that aims to identify a pattern of change, or trend, in a series of observations. The typical goals of trend analysis in cancer surveillance are to determine whether cancer incidence and mortality have increased or decreased over time and to assess the speed with which the increase or decrease has occurred. Cancer incidence and mortality trends are used by public health officials and practitioners for needs assessments, program evaluations, and the development of cancer control strategies. The assessment of cancer trends may be based on the number of cancer cases diagnosed or deaths due to cancer over a select time period; however, to distinguish the impacts of changing population and changing risks, trend analyses for cancer incidence or mortality rates are often preferable. In order to account for variations in population distribution over time, cancer incidence and mortality rates are usually adjusted to the size, sex and age structure of the population. The detection of patterns in these adjusted rates, known as age standardized incidence or mortality rates, permits the identification of cancer trends. While this may be achieved through crudely ‘eye-ballling’ the best fit of rates over time, software that assesses trends and determines the most statistically precise estimate of trends are useful tools in trend analysis calculations. The calculations presented in this report were performed using Joinpoint software, available free through the National Cancer Institute, to accurately identify cancer incidence and
mortality rate trends. Joinpoint software also identifies potential changes in trends at various time points and notes which of these are likely to be the most accurate estimate of the trend.

Although time trend analyses may be used for multiple purposes,² ³ the trend analyses presented in this report are intended to be subsequently used for the estimation of the future number of new cancer cases or deaths. Because recent trends are likely to be the best predictor of cancer rates in consecutive years, extrapolations of the most recent trends are the best predictors of future cancer incidence or mortality. Furthermore, while changes in incidence or mortality rates may be observed using trends established across an entire time course, these broad observations can mask changes in trends within shorter time periods. It is therefore important to evaluate trends in incidence and mortality within various time segments, to identify changes in trends, and to determine the most recent trends in rates and counts. The most recent trend for each cancer site, known as a projection base, can then be used to estimate future incidence or mortality rates and the number of cases or deaths.

Several considerations are necessary when ascertaining time trends that will be extrapolated to calculate future cancer rates. The precision of trend analyses depends on the extent of data available; trend analyses ascertained using data based on a relatively large population, large number of cancer cases diagnosed, and extensive historic data, are more reliable than those based on smaller observations. With fewer data, the precision of the identified trends will decrease and subsequent use of these trends to predict future cancer rates will not be accurate. In addition, although the identification of recent trends is essential for accurate projections of cancer rates, geographic and demographic differences in populations are also important to consider. For example, trend analyses of provincial rates may not be suitable if there are large differences in rates between sub-regions of the province; ascertainment of the trends of rates in smaller geographic areas might lead to more accurate cancer projections. Likewise, potential differences in the age and sex distribution of a population over time should be considered as these demographic characteristics can influence the expected number of cancer cases and deaths. The use of age-specific cancer incidence or mortality rates, rather than age-standardized rates, may produce more accurate future estimates. Also, because projections based on the extrapolation of trends over time assumes that trends in risk behavior and screening or treatment interventions remain the same, these potential influences on future rates should be noted. Assumptions about future cancer incidence and mortality rates based on the extrapolation of recent trends will be more accurate if there are no current interventions that are likely to impact future rates.
In this report, the trends of age standardized cancer incidence and age standardized mortality rates will be explored and change points for incidence and mortality rate trends of 24 cancer sites between 1975 and 2007 will be identified; the methodology used to calculate these trends and decision-making considerations will be discussed. The results of subsequent analyses for the projection of cancer rates will be presented in future Cancer Projection Network reports.
II. Methods

Trend analysis for cancer incidence and mortality is usually based on a regression model relating rates or frequencies to calendar year. Two aspects should be investigated:

a. what is the average change per calendar year determined using an appropriate regression model,

b. is the trend constant over the whole time period investigated using the software “Joinpoint”.

The following sub-sections outline the methods that can be used for the investigation.

1. Regression Models

Trends over time can be investigated by regression analysis, time series analysis and other statistical methods. Regression analysis is the most common and simple method used to analyze trends over time where there is a linear relationship between rate or frequency and calendar year.

A simple linear regression model is often useful to describe the overall change over time. The regression model for the observations: \((x_1, y_1), \ldots, (x_N, y_N)\), where \(x_1 < \ldots < x_N\) represent the time variable, e.g. calendar year and \(y_i, i = 1, 2, \ldots, N\) are the response variable, e.g. the yearly rates, can be written as

\[
E[y_i | x_i] = \beta_0 + \beta_1 x_i
\]

and the coefficient \(\beta_0\) represents the intercept on the y-axis, and the coefficient \(\beta_1\) represents the slope of the line and gives the annual change.

The linear regression model has some underlying assumptions: homogeneous variance, independent observations and normally distributed error term. Thus, the error term is assumed to follow an identical and independent normal distribution (i.i.d.). The least squares method, or maximum likelihood method, can be used for statistical inference and produce the best linear unbiased estimator (BLUE) for parameters in the linear model. The projection can then be obtained by extrapolating the estimated trend from the estimated parameters:

\[
E[y | x] = \hat{\beta}_0 + \hat{\beta}_1 x
\]
given the projection years.

Yearly cancer incidence or mortality data form a time series, and hence the assumptions related to independence and homogeneous variance may not hold. If the response variable follows a normal distribution with different variances (Heteroscedastic), weighted least squares \(^4\) (WLS), or other methods
might be used. Under the weighted least squares method, the coefficients are robust when a weighted sum of squared residuals is minimized and each weight is equal to the reciprocal of the variance. A nonlinear transformation for the response variable may be applied to normalize the data; for instance, the logarithmic transformation.

**Age standardized rates**

If age standardized rates are used, then the variances of these rates are known and homoscedasticity can be tested for using the score test. The weighted least squares method can be used by applying the inverse of the variance of the age standardized rates when variances are not constant. The normality assumption should also be tested, which can be done using a normal probability plot (QQ plot) in the residual analysis. If there are deviations from the diagonal line in the QQ plot, a nonlinear transformation of variables may be applied; for instance, the logarithmic transformation.

The simple linear regression model for logarithmic transformed response is:

\[
E[\log(y_i) | x_i] = \beta_0 + \beta_1 x_i. \tag{3}
\]

If \( \log(y) \) is normalized with constant variance, the least squares method can be applied to estimate the parameters. However, when the age standardized rates are not homoscedastic, the weighted least squares method should be used to estimate the parameters. Again, projections can be estimated by

\[
E[y | x] = e^{\hat{\beta}_0 + \hat{\beta}_1 x} \tag{4}
\]
given the projection years. The annual percentage change of \( y \) is estimated by \( e^{\hat{\beta}_1} - 1 \).

**Frequencies**

When the response variable, \( y \), is the count of cancer cases or deaths and follows the Poisson distribution, the constant variance assumption is broken because the variance is the same as the mean (count) which may be different across years. In this situation, the weighted least squares method is used to estimate the parameters. For population based research, the cancer counts are related to the size of the population; therefore, population size should be accounted for in the model using an ‘offset’ term. The log linear regression model is then:

\[
E[\log(y_i) | x_i] = \beta_0 + \beta_1 x_i + \ln(p_i) \quad \text{or} \quad E[\log\left( \frac{y_i}{p_i} \right) | x_i] = \beta_0 + \beta_1 x_i \tag{5}
\]
where \( p_i \) is the population in the \( i \)-th year. Compared to the age standardized rate as response, \( \frac{y_i}{p_i} \) is the crude rate of cancer incidence or mortality and does not take into account the changing population structure.

2. Joinpoint Regression Model

The methods presented in the previous section do not allow for changes in the trend over time. For some cancer sites, changing exposure to risk factors and/or the introduction of new screening programs or other interventions may affect the trend in rates and frequencies. Sudden changes in population structure, such as new patterns of migration, may also result in changes in the trends in the actual numbers of cancers. Thus, in modeling trends over time, it is important to be able to detect when statistically significant changes in the trend occur. Joinpoint analysis is widely applied to detect these changes points (joinpoints) and determine the trends between joinpoints.\(^5,6\)

The joinpoint regression model for the observations: \((x_1, y_1), \ldots, (x_N, y_N)\), where \( x_1 < \ldots < x_N \) represent the time variable, e.g. calendar year and \( y_i, i = 1, 2, \ldots, N \) are the response variable, e.g. the annual age standardized rates or frequencies, can be written as

\[
E[y_i | x_i] = \beta_0 + \beta_1 x_i + \gamma_1 (x_i - \tau_1)^+ + \cdots + \gamma_n (x_i - \tau_n)^+ + \epsilon_i
\]

where \( \beta_0, \beta_1, \gamma_1, \cdots, \gamma_n \) are regression coefficients and the \( \tau_k, k = 1, 2, \ldots, n, n < N \), is the \( k \)-th unknown joinpoint in which

\[
(x_i - \tau_k)^+ = (x_i - \tau_k) \text{ if } (x_i - \tau_k) > 0; \quad (x_i - \tau_k)^+ = 0, \text{ otherwise,}
\]

This model assumes a linear trend between joinpoints and continuity at the joinpoints. The joinpoint method is also known as piecewise regression, segmented regression,\(^7\) broken line regression, or multi-phase regression with the continuity constraint.

The joinpoint regression model has the same underlying assumptions as simple regression. Noting that the homogeneous variance and independence assumptions are usually not valid for time series data, heteroscedastic variances are assumed and the weighted least square method is employed for the inference
and computation of the Joinpoint linear regression model. The underlying distribution for response variable $y$ or its transformation can be assumed as:

(i) $y_i \sim N(\mu_i, \sigma_i^2)$ for homoscedastic rates or log-rates

(ii) $y_i \sim N(\mu_i, \sigma_i^2)$ for heteroscedastic rates or log-rates

(iii) $y_i \sim Pois(\mu_i)$ for counts without adjustment (offset) for population size

(iv) $y_i \sim Pois\left(\frac{\mu_i}{p_i}\right)$ for counts with adjustment (offset) for population size

If the year(s) when changes in the trend occur (joinpoints) are known, then linear regression techniques can be used to estimate the regression parameters. However, in most instances, the exact years for the joinpoints are unobservable. The challenge in cancer trend analysis is to determine the locations of the joinpoints if they exist; and to determine the optimal number of joinpoints for the most appropriate model.

**Fitting the model**

There are three major decisions in any joinpoint analysis:

1. The form of the mean function (Data distribution: Normal or Poisson; Equation: linear or log linear),
2. The location of the joinpoints given the number of joinpoints, and
3. The optimal joinpoint model.

The first step is determined by the form of the data, as described above. The next step in fitting the model is to determine the range of the number of joinpoints to be tested; usually between 0 and 4. Then for each given number of joinpoints the location of the joinpoints is determined. To determine the location of the joinpoints, either the grid search method or Hudson’s method can be applied. The grid search method creates a "grid" of all possible locations for joinpoints specified by the settings, and tests the sum square of errors (SSE) at each one to find the best possible fit. Hudson's method does a continuous testing between observed values to find the best model. In Hudson's method, each fitted sub-model has a local least square, and the overall least squares of the model is found when a complete curve to be fitted consists of two or more sub-models.
The third step is to find the optimal model, i.e. the optimal number of joinpoints, and the optimal locations of related joinpoints. There are two major classifications of methods to determine the optimal model: hypothesis testing and information criteria. Stepwise selection procedures base on the hypothesis testing approach sequentially apply classical $F$-tests to test whether some of the regression parameters are zero. The Bayesian information criterion (BIC) is based on the sum of two parts: the goodness-of-fit of a candidate model and the penalty for model complexity.

Kim et al (2000) proposed a series of permutation tests to determine the best number of change-points in segmented line regression. In choosing the better model between the one with $k$ change-points and the alternative with $l$ change-points, Kim et al proposed the Monte Carlo simulation approach for permutation tests based on the $F$-statistics:

$$F_{(0)} = \frac{RSS(l)}{RSS(k)}$$

obtained from the residual sum of squares: $RSS(k)$ and $RSS(l)$, for the model $k$ and $l$, where the degree of freedom of the $F$-distribution are $k$ and $l$, respectively. Then using Monte Carlo simulation approach to calculate the frequency that the event $[F_{(i)} \geq F_{(0)}]$ happened, for the $F$-statistics: $F_{(i)}$, $i = 1, 2, \ldots, N - 1$:

$$p = \frac{\#[F_{(i)} \geq F_{(0)}]}{N}$$

The hypothesis test $H_0$: model $k$ vs $H_1$: model $l$, is implemented by comparing the probability $p$ with the adjusted significant level $\alpha'$, which is defined as the Bonferroni significant value given the significant level $\alpha$. For example, when $\alpha = 0.05$, $k = 1$ from previous permutation test: $k = 0$ and $l = 5$, and $l = 5, 4, 3, \text{or} 2$, then $\alpha' = \alpha / 4 = 0.0125$.

The sequential method to determine the optimal model using the permutation test is described below:

a. Set $k_1$ — the maximum number of joinpoints, $k_0$ — the minimum number of joinpoints, $0 \leq k_0 < k_1$, and calculate $\alpha' = \frac{\alpha}{k_1 - k_0}$ for a given significant level $\alpha$;

b. Test $H_0$: $k_0$ is the best model, against $H_1$: $k_1$ is the best model;

c. Calculate $F_{(0)}$, $F_{(i)}$, $i = 1, 2, \ldots, N - 1$, and Monte Carlo $p$-value by (8) and (9);

i. If $p > \alpha'$, do not reject $H_0$ and set $k_1 = k_1 - 1$. 

11
(ii) If \( p \leq \alpha' \), then \( H_0 \) is rejected; Set \( k_0 = k_0 + 1 \), and calculate new \( \alpha' = \frac{\alpha}{k_1 - k_0} \);

d. Go back to do step b and c until \( k_1 = k_0 + 1 \).

Usually, the optimal model determined by BIC can be easily achieved by two processes as follows:

a. Calculate the BIC for each joinpoint model,

b. Select the joinpoint model with minimum BIC as the optimal model.

3. Joinpoint Regression Software

The input data for the joinpoint system include either the annual age standardized incidence or mortality rates or the annual incident or death counts and, if applicable, the corresponding population data. If the data are heteroscedastic and the weighted least square method is to be used, the variance of each rate must also be input. Only one independent variable is allowed, and in trend analyses for cancer rates and counts this is assumed to be calendar year which must be input.

The following parameters must also be entered:

(1) *The form of the equation*, which is dependent on the underlying distribution of the response variable (see section II.1.).

a. *Homoscedastic errors*: These simple regression models can be used where it is known that the response variables are homoscedastic or where the variances of the response variables are not known.

i. *Simple linear regression*: \( y_i = \beta_0 + \beta_1 x_i \), where \( y_i \) is age standardized rate, \( x_i \) is calendar year, the \( \beta \)'s are the regression parameter. The linearity assumption should be evaluated and the regression coefficients are estimated using the least squares method. The \( \beta_1 \) is the estimate of the slope and is equal to the annual change in rate;

ii. *Logarithmic transformation*: \( \log(y_i) = \beta_0 + \beta_1 x_i \);  This function may be used to normalize the response variable, and is frequently used for age standardized rates; or this function may be used for the Poisson model as the link function, and is frequently used for cancer cases. Again, the goodness of fit of this model must be evaluated. In this model the annual percentage change (APC) is \(( e^{\beta_1} - 1 ) \ast 100 \).
b. *Heteroscedastic errors*: These methods are used where there is heteroscedasticity, and the variances are known or can be estimated.

i. *Normally distributed response variables*: This method is used where the response variable \((y)\) can be assumed to be normally distributed with heteroscedastic variances. This is usually the situation with age standardized rates. A weighted regression analysis is then appropriate using the inverse of the variance \((\nu)\) of the age standardized rates as the weight for the linear regression model, and the weight \((w = (y^2)/\nu))\) for the log-transformed data, by delta method. The annual change (linear model) and the annual percentage change (log linear model) are estimated as in the unweighted analysis;

ii. *Poisson model using count*: This method is used where the response variable is the count, and there is no adjustment for changes in population size and distribution. It is assumed that the count has a Poisson distribution with mean and variance estimated by the count \((c)\). Thus, the weight used for the identity link is the inverse of the count \((1/c)\) or, for the log link, the count \((c)\). Again, joinpoint uses the weighted least squares method for the model fitting, as opposed to maximum likelihood. In the situation where there are annual counts of zero, the counts are adjusted by the addition of 0.5 to allow fitting of the model.

iii. *Poisson model using rate*: This model is usually used where the response variable is a crude rate \((\text{count} \div \text{population})\). It is assumed that the count has a Poisson distribution and that the population is included in the model as an offset. The weight can be defined as \(w = p^2/c\) for identity link function or \(w = c\) for log link function. If there are annual counts of zero, the counts are adjusted as in (ii) under (b).

(2) *Method to test for the best fit model*. There are two methods to select the best joinpoint model in Joinpoint software. One method is the permutation test using a series of simulation procedures. The other method uses Bayesian Information Criterion (BIC).

a. *Permutation test*: If the permutation test is used, the overall significance level for the permutation test and the number of randomly permuted data sets for the permutation
test must be set. A significance level ($\alpha$) 0.05 is usually used. The second parameter is to control the number of permutations; the larger the number of permutations, the more consistent the $p$-values obtained. The default number of permutations is 4499,\(^1\) which is adequate for most situations. In applying the significance tests, the Bonferroni adjustment is used to adjust the significance level for multiple comparisons. To reach a conclusion, Joinpoint performs $k_1 - k_0$ tests, and adjusts the significance level to $\alpha/(k_1 - k_0)$;

b. *Bayesian Information Criterion*: The value of BIC is the log likelihood value with penalty of extra parameters. The model with the minimum BIC value is the best Joinpoint model.

(3) The number of joinpoints. The minimum and maximum number of joinpoints that will be tested: In general the minimum number would be taken as zero, i.e. a uniform linear relation, and the maximum number would depend on the number of years of available data and the methods used to find the location of the joinpoints. The Grid search method allows a minimum of zero and a maximum of nine joinpoints, while the Hudson’s method allows a minimum of zero and a maximum of four. Other considerations are:

a. *Number of observations between joinpoints*. The suggested minimum number of observations between two joinpoints is five. This is to reduce over-fitting of the model;

b. *Number of observations in the last segment*. If the trend analysis is to be used for projections, the suggested minimum here is 6 to ensure robust estimates of the slope.

(4) Searching method. There are two fitting methods to search for the locations of joinpoints: The Grid method and Hudson’s method. The Grid search only tests a discrete number of locations, while Hudson’s method does continuous testing between observed values to find the best model, which allows the joinpoints to occur anywhere between observations.

a. *Number of grid points* (grid point method only). The number of grid points (from zero to nine) to place between adjacent observations must be specified. If this value is zero, the grid will be the observations. If this value is one, the grid will be the observations plus the midpoints between observations, *etc.* It is recommended that 0 grid points be used;
(5) **Defining the ‘by’ variable (optional).** The user can define the ‘by’ variable to perform separate analyses for each by-group at the same time. This option saves time when analyzing sub-groups, where the model and other parameters are the same for all sub-groups; for example, site and sex. The results of each by-group will be displayed sequentially. This option provides a comparison among trends in all the by-groups.

Several forms of output are available, including data, model fitting and other statistics and graphs, after running a model in the Joinpoint software. These are detailed in the following section.
III. Results and Conclusions

1. Data Preparation

Cancer incidence/mortality data are from the Alberta Cancer Registry, 1975 – 2007. During the study period, three different revisions of the International Classification of Diseases for Oncology (ICDO) and International Classification of Diseases (ICD) were used. To improve comparability, cancer incidence was converted into ICDO-3 and cancer mortality was converted into ICD-10. To be consistent with CCS publications, cancer incidence/ mortality data were grouped into CCS cancer site definitions. Alberta population data was from Statistics Canada, which are based on the official censuses.

Each cancer site was grouped by sex, calendar year, and 5-year age group (from 0-4 to 80-85 and 85+ years). Age standardized incidence rates and age standardized mortality rates were computed per 100,000 population. Rates are standardized using the 1991 Canadian standard population. A SAS program was applied to calculate the age standardized rates, but SEER Stat may also be used.

2. Parameter Setting in Joinpoint Software

In general, if there is no zero value in rates for a cancer site, the log-transformation is used to analyze the trends of age standardized rates. If there is a zero value in rates for a cancer site, the logarithmic-transformation is not suitable; the original ASRs should be used assuming constant variance. Alternatively, the log-linear regression model under Poisson distribution can be used for crude rates (yearly counts / population), adjusting the counts by the addition of a small constant, such as 0.5. In Alberta, mortality for female larynx, liver and thyroid, and male testis and thyroid had zero counts for one or more years and this model was used for trend analysis.

A flow chart of the parameter settings in Joinpoint regression software is shown in Figure 1. For analysis of Alberta data, the following settings for the parameters were used:

a. Model: ln(y) = bx. The log transformation of the age standardized incidence rate or age standardized mortality rate was used (see section II.3.(1).a.ii, p12).

b. Heteroscedastic errors option: The standard error of each age standardized rate was input.

c. By Variable: cancer site and sex. Site and sex specific joinpoint trend analysis was conducted. This allows more than one site to be analyzed at one time.
d. Number of Joinpoints: min = 0 and max = 5. The allowed maximum number of joinpoints is 5 over 33 years as at least 6 years is required for each segment.

e. Method: Grid Search. The Hudson’s method is very time-consuming (see section II.3.(4), p14)

f. Number of Observations: Minimum number of observations from a joinpoint to either end of the data is 6; Minimum number of observations between two joinpoints = 6;

h. Model Selection Method: Permutation Test. Significance level is 0.05 as common and the number of permutation is 4499 as default (see section II.3.(2).a, p13).

Figure 1. Flow chart of the parameter settings and decision tree
3. Outputs
Female colorectal cancer incidence was used to illustrate how the joinpoint analysis and Joinpoint software work. Parameter settings are illustrated in Table 1. The results from Joinpoint can be accessed through five window tabs: Permutation, Model estimates, Trends and Graphs.

Table 1. Parameter settings for female colorectal cancer incidence

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>( \ln(y) = bx )</td>
</tr>
<tr>
<td>Dependent variable</td>
<td>Age standardized rate</td>
</tr>
<tr>
<td>Independent variable</td>
<td>year</td>
</tr>
<tr>
<td>Heteroscedastic errors option</td>
<td>Input Standard Error of Dependent Variable</td>
</tr>
<tr>
<td>By variable</td>
<td>Cancer site and sex</td>
</tr>
<tr>
<td>Number of Joinpoints</td>
<td>min = 0 and max=5</td>
</tr>
<tr>
<td>Method</td>
<td>Grid Search</td>
</tr>
<tr>
<td>Minimum number of observations from a joinpoint to either end of the data</td>
<td>6</td>
</tr>
<tr>
<td>Minimum number of observations between two joinpoints</td>
<td>6</td>
</tr>
<tr>
<td>Model Selection Method</td>
<td>Permutation Test</td>
</tr>
<tr>
<td>Significance level</td>
<td>0.05</td>
</tr>
<tr>
<td>Number of permutations</td>
<td>4499</td>
</tr>
</tbody>
</table>

The Permutation Tests tab shows the testing process to determine the optimal model. The testing process is that outlined in section II.2. There are 6 possible joinpoint regression models in the analysis. Model 0 (0 joinpoints) with model 5 (5 joinpoints) is the first test. Using the Bonferroni adjustment, the significance level\(^1\) for testing is \( \alpha' = \alpha/5 (= 0.01) \). The \( p \)-value is less than the adjusted significance level, indicating that the model with 5 joinpoints is the preferred model; thus, the next test is model 1 against model 5 using an adjusted significance of \( \alpha' = \alpha/4 (= 0.0125) \). In this instance, \( p > 0.0125 \), so the next test would be model 1 against model 4, again using a significance level of \( \alpha' = \alpha/4 \), etc. The final conclusion from this series of testing is that, for female colorectal incidence in Alberta, the model with 1 joinpoint is the best model. The procedures are listed in Table 2 (* is the selected model for each level of testing).
The **Graph** tab shows the graph of trend. The APC values of the trend for each segment are presented in the top-right of Figure 2. For female colorectal cancer between 1975 and 1995 there was a significant downward trend of –0.90%; between 1995 and 2007 there was a significant upward trend of 0.85%.

The **Data** tab shows the observed values, model estimated values and the joinpoint location, in this case, 1995 (Table 3).

The **Model Estimates** tab shows the summary of model estimates including the mean square error and degrees of freedom, the joinpoints with their 95% confidence intervals, the intercept and slope, and their confidence intervals for each segment (Table 4).

The **Trends** tab shows the APC and confidence intervals for each segment and Average Annual Percentage Change (AAPC) for specified numbers of the most frequent years. The default is 5 and 10 years, but these can be set in the set up phase (Table 5).

Table 2. Permutation testing procedure – Alberta Female Colorectal Cancer

<table>
<thead>
<tr>
<th>Test For Number of Joinpoints</th>
<th>Cohort</th>
<th>Test Number</th>
<th>Null Hypothesis</th>
<th>Alternate Hypothesis</th>
<th>Numerator Degrees of Freedom</th>
<th>Denominator Degrees of Freedom</th>
<th>Number of Permutations</th>
<th>P-Value</th>
<th>Significance Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>H1</td>
<td>0 Joinpoint(s)</td>
<td>5 Joinpoint(s) *</td>
<td>10</td>
<td>21</td>
<td>4500</td>
<td>0.0006667</td>
<td>0.010000</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>H2</td>
<td>1 Joinpoint(s) *</td>
<td>5 Joinpoint(s)</td>
<td>8</td>
<td>21</td>
<td>4500</td>
<td>0.0192222</td>
<td>0.012500</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>H3</td>
<td>1 Joinpoint(s)</td>
<td>4 Joinpoint(s)</td>
<td>6</td>
<td>23</td>
<td>4500</td>
<td>0.0406667</td>
<td>0.012500</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>H4</td>
<td>1 Joinpoint(s)</td>
<td>3 Joinpoint(s)</td>
<td>4</td>
<td>26</td>
<td>4500</td>
<td>0.0106667</td>
<td>0.012500</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>H5</td>
<td>1 Joinpoint(s)</td>
<td>2 Joinpoint(s)</td>
<td>2</td>
<td>27</td>
<td>4500</td>
<td>0.0135356</td>
<td>0.012500</td>
<td></td>
</tr>
</tbody>
</table>

Final Selected Model: 1 Joinpoint(s)
Figure 2. Joinpoint regression lines for female colorectal cancer incidence in Alberta.

Table 3. Output datasets from Joinpoint for female colorectal cancer incidence in Alberta.

<table>
<thead>
<tr>
<th>X Value</th>
<th>Observed Y Value</th>
<th>Modelled Y Value</th>
<th>Joinpoint Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>1975</td>
<td>42.01</td>
<td>43.42</td>
<td></td>
</tr>
<tr>
<td>1976</td>
<td>39.55</td>
<td>43.03</td>
<td></td>
</tr>
<tr>
<td>1977</td>
<td>42.59</td>
<td>42.65</td>
<td></td>
</tr>
<tr>
<td>1978</td>
<td>39.20</td>
<td>42.27</td>
<td></td>
</tr>
<tr>
<td>1979</td>
<td>44.42</td>
<td>41.89</td>
<td></td>
</tr>
<tr>
<td>1980</td>
<td>40.92</td>
<td>41.51</td>
<td></td>
</tr>
<tr>
<td>1981</td>
<td>39.65</td>
<td>41.14</td>
<td></td>
</tr>
<tr>
<td>1982</td>
<td>41.93</td>
<td>40.77</td>
<td></td>
</tr>
<tr>
<td>1983</td>
<td>43.70</td>
<td>40.41</td>
<td></td>
</tr>
<tr>
<td>1984</td>
<td>40.08</td>
<td>40.04</td>
<td></td>
</tr>
<tr>
<td>1985</td>
<td>45.94</td>
<td>39.69</td>
<td></td>
</tr>
<tr>
<td>1986</td>
<td>40.94</td>
<td>39.33</td>
<td></td>
</tr>
<tr>
<td>1987</td>
<td>38.21</td>
<td>38.38</td>
<td></td>
</tr>
<tr>
<td>1988</td>
<td>35.44</td>
<td>38.63</td>
<td></td>
</tr>
<tr>
<td>1989</td>
<td>38.80</td>
<td>38.28</td>
<td></td>
</tr>
<tr>
<td>1990</td>
<td>38.13</td>
<td>37.94</td>
<td></td>
</tr>
<tr>
<td>1991</td>
<td>36.33</td>
<td>37.60</td>
<td></td>
</tr>
<tr>
<td>1992</td>
<td>35.94</td>
<td>37.27</td>
<td></td>
</tr>
<tr>
<td>1993</td>
<td>36.12</td>
<td>36.93</td>
<td></td>
</tr>
<tr>
<td>1994</td>
<td>36.88</td>
<td>36.60</td>
<td></td>
</tr>
<tr>
<td>1995</td>
<td>35.70</td>
<td>36.27</td>
<td>Joinpoint 1</td>
</tr>
<tr>
<td>1996</td>
<td>35.77</td>
<td>36.58</td>
<td></td>
</tr>
<tr>
<td>1997</td>
<td>37.27</td>
<td>36.89</td>
<td></td>
</tr>
<tr>
<td>1998</td>
<td>37.81</td>
<td>37.21</td>
<td></td>
</tr>
<tr>
<td>1999</td>
<td>35.11</td>
<td>37.52</td>
<td></td>
</tr>
</tbody>
</table>
Table 4. The model estimates information – Alberta Female Colorectal Cancer

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Number of Joinpoints</th>
<th>Number of Observations</th>
<th>Number of Parameters</th>
<th>Degrees of Freedom</th>
<th>Sum of Squared Errors</th>
<th>Mean Squared Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>1</td>
<td>33</td>
<td>4</td>
<td>29</td>
<td>30.58836</td>
<td>1.05477</td>
</tr>
</tbody>
</table>

Estimated Joinpoints

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Joinpoint</th>
<th>Estimate</th>
<th>Lower CI</th>
<th>Upper CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>1</td>
<td>1995</td>
<td>1990</td>
<td>1999</td>
</tr>
</tbody>
</table>

Estimated Regression Coefficients (Beta)

### Standard Parameterization

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Parameter</th>
<th>Parameter Estimate</th>
<th>Standard Error</th>
<th>Z</th>
<th>Prob &gt;</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>Intercept 1</td>
<td>21.532542</td>
<td>4.243383</td>
<td>5.074381</td>
<td>0.000023</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>Slope 1</td>
<td>-0.008993</td>
<td>0.002137</td>
<td>-4.207683</td>
<td>0.000240</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>Slope 2 - Slope 1</td>
<td>0.017471</td>
<td>0.004189</td>
<td>4.171079</td>
<td>0.000265</td>
<td></td>
</tr>
</tbody>
</table>

### General Parameterization

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Parameter</th>
<th>Parameter Estimate</th>
<th>Standard Error</th>
<th>Z</th>
<th>Prob &gt;</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>Intercept 1</td>
<td>21.532542</td>
<td>4.243383</td>
<td>5.074381</td>
<td>0.000023</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>Intercept 2</td>
<td>-13.321638</td>
<td>7.211194</td>
<td>-1.847356</td>
<td>0.075287</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>Slope 1</td>
<td>-0.008993</td>
<td>0.002137</td>
<td>-4.207683</td>
<td>0.000240</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>Slope 2</td>
<td>0.008478</td>
<td>0.003602</td>
<td>2.353446</td>
<td>0.025853</td>
<td></td>
</tr>
</tbody>
</table>

Table 5. The trends tab with the APC and AAPC in each trend – Alberta Female Colorectal Cancer

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Joinpoint</th>
<th>Estimate</th>
<th>Lower CI</th>
<th>Upper CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>1</td>
<td>1995</td>
<td>1990</td>
<td>1999</td>
</tr>
</tbody>
</table>

### Annual Percent Change (APC)

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Segment</th>
<th>Lower Endpoint</th>
<th>Upper Endpoint</th>
<th>APC</th>
<th>Lower CI</th>
<th>Upper CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>1</td>
<td>1975</td>
<td>1995</td>
<td>-0.9*</td>
<td>-1.3</td>
<td>-0.5</td>
</tr>
<tr>
<td>F</td>
<td>2</td>
<td>1995</td>
<td>2007</td>
<td>0.9*</td>
<td>0.1</td>
<td>1.6</td>
</tr>
</tbody>
</table>

* The Annual Percent Change (APC) is statistically significant from zero.

### Average Annual Percent Change (AAPC)

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Range</th>
<th>Lower Endpoint</th>
<th>Upper Endpoint</th>
<th>AAPC</th>
<th>Lower CI</th>
<th>Upper CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>Last 5 Obs.</td>
<td>2003</td>
<td>2007</td>
<td>0.9*</td>
<td>0.1</td>
<td>1.6</td>
</tr>
<tr>
<td>F</td>
<td>Last 10 Obs.</td>
<td>1998</td>
<td>2007</td>
<td>0.9*</td>
<td>0.1</td>
<td>1.6</td>
</tr>
</tbody>
</table>

* The Average Annual Percent Change (AAPC) is statistically significant from zero.
4. Summary Results

The results are summarized here for each cancer site. The corresponding figures can be found in Appendix, Fig. A1 – A24.

(1) All Cancers: There was one joinpoint at 2002 for female incidence: the rate increased by 0.95% per year from 1975 to 2002 and stabilized from 2002 to 2007. For male incidence, there was one joinpoint at 2001: the rate increased by 1.22% per year from 1975 to 2001 and then decreased by 1.15% from 2001 to 2007. For female mortality, there was one joinpoint at 1996: the rates increased by 0.67% per year from 1975 to 1996 and decreased by 1.01% per year from 1996 to 2007. For male mortality, there was one joinpoint at 1991: the rates increased by 0.70% per year from 1975 to 1991 and decreased by 1.00% per year from 1991 to 2007.

(2) Colorectal cancer: There was one joinpoint at 1975 for female incidence: the rate decreased by 0.90% per year from 1975 to 1995 and increased by 0.85% per year from 1995 to 2007. For male incidence, there was one joinpoint at 1983: the rate increased by 2.59% per year from 1975 to 1983 and then slowly increased by 0.35% from 1983 to 2007. For female mortality, ASRs decreased by 1.38% per year overall from 1975 to 2007 and there was no significant joinpoint. For male mortality, there was one joinpoint at 1982: the rate stabilized from 1975 to 1982 and decreased by 0.78% per year from 1982 to 2007.

(3) Lung cancer: For female incidence, there were two joinpoints at 1984 and 1996: the rate increased by 8.27% per year from 1975 to 1984, increased by 3.51% per year from 1984 to 1996 and slowly increased by 1.52% from 1996 to 2007. For male incidence, there was one joinpoint at 1987: the rate increased by 2.27% per year from 1975 to 1987 and then decreased by 1.24% per year from 1987 to 2007. For female mortality, there was one joinpoint at 1993: ASRs increased by 5.48% per year from 1975 to 1993 and slowly increased by 1.25% per year from 1993 to 2007. For male mortality, there was one joinpoint at 1988: ASRs increased by 1.75% per year from 1975 to 1988 and decreased by 1.49% per year from 1988 to 2007.

(4) Breast cancer of female: For incidence, there was one joinpoint at 1999: the rate increased by 1.34% per year from 1975 to 1999 and stabilized from 1999 to 2007. For mortality, there was one joinpoint at 1991: ASRs increased by 0.75% per year from 1975 to 1991 and decreased by 2.90% per year from 1991 to 2007.
(5) Prostate cancer: For incidence, there was one joinpoint at 2001: the rate increased by 3.60% per year from 1975 to 2001 and decreased by 3.96% from 2001 to 2007. For mortality, there was one joinpoint at 1995: ASRs increased by 0.80% per year from 1975 to 1995 and decreased by 3.12% per year from 1995 to 2007.

(6) Oral cancer: ASRs of female incidence were stable over 1975 to 2007. For male incidence, the overall ASRs decreased 2.56% per year from 1975 to 2007. For mortality, both female ASRs and male ASRs were stable between 1975 and 2007.

(7) Bladder cancer: ASRs of female incidence increased 0.60% per year from 1975 to 2007. For male incidence, there was one joinpoint at 1982: the rate increased by 3.08% per year from 1975 to 1982 and stabilized from 1982 to 2007. For mortality, both female ASRs and male ASRs were stable between 1975 and 2007.

(8) Brain cancer: The ASRs of both female and male brain cancer incidence were stable from 1975 to 2007. Over the same period, the mortality ASRs were stable for both females and males.

(9) Cervical cancer: ASRs of incidence decreased 1.43% per year from 1975 to 2007 while mortality decreased 2.16% per year from 1975 to 2007.

(10) Esophageal cancer: ASRs of incidence increased 1.45% and 3.07% per year from 1975 to 2007 for females and males, respectively. Female mortality stabilized overall while male mortality increased 2.12% per year from 1975 to 2007.

(11) Hodgkin lymphoma: ASRs of female incidence increased 1.00% per year from 1975 to 2007 while the regression line of male incident ASRs were stable over the same period. In contrast, for female mortality, there was one joinpoint at 1990: the rate decreased by 10.42% per year from 1975 to 1990 and then stabilized from 1990 to 2007. The ASRs of male mortality decreased 3.71% over 1975 to 2007.

(12) Kidney cancer: For female incident ASRs, there was one joinpoint at 1987: ASRs increased 4.81% per year from 1975 to 1987 and increased 0.90% per year from 1987 to 2007. Male incident ASRs also
had one joinpoint at 1991: ASRs increased 3.58% per year and then stabilized from 1991 to 2007. In contract, for mortality, the overall ASRs for both females and males were stable from 1975 to 2007.

(13) Larynx cancer: ASRs of female incidence were stable over the period 1975-2007. There were zero case in some years for female mortality and the Poisson model was applied to the trend analysis for crude rates. The crude rates increased 2.04% per year from 1975 to 2007. For males, ASRs of incidence and mortality decreased by 1.54% per year from 1975 to 2007.

(14) Leukemia: ASRs of incidence increased 0.56% and 0.38% per year from 1975 to 2007 for females and males, respectively. In contract, ASRs for mortality decreased 1.01% and 0.93% per year from 1975 to 2007 for females and males, respectively.

(15) Liver cancer: ASRs of female incidence increased 2.52% per year from 1975 to 2007. There were zero cases in some years for female mortality and the Poisson model was applied to the trend analysis using crude rates. The crude rates increased 4.52% per year from 1975 to 2007. For males, ASRs increased 3.57% and 2.92% per year from 1975 to 2007 for incidence and mortality, respectively.

(16) Melanoma of skin: For female incidence, there was one joinpoint at 1999: ASRs increased 2.96% per year from 1975 to 1999 and then stabilized from 1999 to 2007. For male incidence, there was one joinpoint at 1995: ASRs increased 5.64% per year from 1975 to 1995 and then stabilized from 1995 to 2007. For mortality, ASRs stabilized over time for females; in contract, ASRs increased 1.73% per year for males from 1975 to 2007.

(17) Multiple myeloma: Overall, ASRs of incidence and mortality were stable from 1975 to 2007 for both males and females, respectively.

(18) Non-Hodgkin’s Lymphoma: ASRs of incidence increased 1.77% and 1.84% per year from 1975 to 2007 for females and males, respectively. Female mortality rates were stable from 1975 to 2007 while male mortality rates increased 0.61% per year from 1975 to 2007.

(19) Ovary cancer: There was one joinpoint at 1993: ASRs stabilized from 1975 to 1993, and decreased 3.62% per year from 1993 to 2007. ASRs of mortality declined 1.02% per year from 1975 to 2007.
(20) Pancreas cancer: Overall, ASRs of female incidence increased 0.49% per year from 1975 to 2007. In contrast, ASRs decreased 0.65% per year from 1975 to 2007 for male incidence. Female mortality rates were stable and male mortality rates decreased 1.09% per year from 1975 to 2007.

(21) Stomach cancer: ASRs of incidence decreased 2.11% and 2.50% per year from 1975 to 2007 for females and males, respectively. For mortality, ASRs decreased 2.64% and 2.86% per year from 1975 to 2007 for females and males, respectively.

(22) Testis cancer: ASRs of incidence increased 1.83% per year from 1975 to 2007. There were zero cases in some years for mortality and the Poisson model was applied to the trend analysis for crude rates. The crude rates decreased 9.17% per year from 1975 to 1988 and stabilized between 1988 and 2007.

(23) Thyroid cancer: For female incidence, there was one joinpoint at 1995: ASRs increased 1.96% per year from 1975 to 1995, and then rapidly increased 7.56% per year from 1995 to 2007. For male incidence, ASRs increased 3.41% per year from 1975 to 2007. The crude rates of female mortality were stable from 1975 to 2007 and the crude rates of male mortality were stable from 1975 to 2007.

(24) Body of uterus cancer: Overall, the ASRs of incidence were stable from 1995 to 2007. ASRs of mortality decreased 0.58% per year from 1975 to 2007.

5. Results for Special Consideration

The impact of changing parameter settings and different models in Joinpoint software were investigated; the results are as follows:

(1) Original rates ($y$) vs log-transformed rates ($\log(y)$):

For trend analysis of age standardized rates, if the data are normally distributed, no transformation is required; if not, the log-transformation can be applied to normalize the data. The APC of the trend in original rates is usually not significant, while that of the trend in log-transformed rates is usually significant. For example, for female breast cancer incidence, both regressions found the optimal model with one joinpoint at year 1999; however, APCs of the trends in the two segments are not significant for original rates, while the APCs are significant for log-transformed rates. Residual analysis (section III.5.10, p30) can be used to determine the better fit.
Homogeneous variance vs heterogeneous variance:
When ASRs are assumed to have constant (homogeneous) variance, both least square estimators and maximum likelihood estimators are equivalent. When ASRs are assumed to have different estimated variances, the weighted least square method is applied in which the weights are defined as the inverse of estimated variance of ASRs. In general, the identified joinpoints are the same between the two options; however, the estimated APC(s) may be slightly different from each other. For example, for female esophageal cancer incidence, the APC is 1.70% from 1975 to 2007 using constant variance and 1.45% using weighted least square method. For female breast cancer, both options select the joinpoint at year 1995: APCs are –0.85% (in 1975 – 1995) and 0.84% (in 1995 – 2007) using constant variance; but, APCs are –0.90% (in 1975 – 1995) and 0.85% (in 1995 – 2007) using weighted least square method. Residual analysis (section III.5.10, p30) can be used to determine the better fit.

Permutation test vs BIC method:
In general, the computation time using BIC method is at least five times shorter than that for the permutation test. In the trend analysis of age standardized rates of incidence, the same optimal model was selected using BIC as the permutation test for most cancer sites, with exceptions as follows:

1. female colorectal cancer (Figure 3),
   a. BIC – two joinpoints were identified at 1985 and 1991,
   b. Permutation test – one joinpoint was selected at 1995
2. female Hodgkin’s lymphoma,
   a. BIC – two joinpoints were identified at 1987 and 1994,
   b. Permutation test – no joinpoint was identified.
3. male esophageal cancer,
   a. BIC – one joinpoint was identified at 1990,
   b. Permutation test – no joinpoint was identified
4. male larynx cancer,
   a. BIC – one joinpoint was identified at 1988
   b. Permutation test – no joinpoint was identified
5. male kidney cancer,
   a. BIC method – two joinpoints were identified at 1980 and 1986,
   b. Permutation test – one joinpoint was identified at 1991
6. male thyroid cancer,
   a. BIC – one joinpoint was identified at 1997,
   b. Permutation test – no joinpoint was identified

In general, the permutation test is more conservative than the BIC. Although for some sites where there is a different number of joinpoints the slope of the last segment is similar (Figure 3), this is not the case for all such sites. Since, in general, the most parsimonious model is preferred, the use of the Permutation test is recommended even though it takes more computer time. It may be worthwhile, however, to check the model where the data may warrant it.

Figure 3. Optimal joinpoint trends for ASRs of Alberta female colorectal cancer incidence using BIC method (red lines) and permutation test method (blue lines)

(4) Grid search vs Hudson’s method:
The major difference between the Grid search method and Hudson’s method is in computation time; Hudson’s method was ten times slower than the Grid search method. The locations of the identified joinpoints are similar. If the joinpoints for the Grid method are constrained to be at the calendar year, the joinpoints identified by Hudson’s method are within one calendar year of this joinpoint. The difference in the resultant APCs for the two methods is small. For example, for male lung cancer incidence, one joinpoint was identified using the Grid search method at 1987 and the APC was −1.24% between 1987 and 2007; the joinpoint identified using Hudson’s method was 1987.3 and the APC was −1.27% between 1987.3 and 2007.
(5) ASR vs crude rate:
Both the ASRs and the crude rates reflect the overall rates; however, the ASR adjusts for changes in population size and age structure while the crude rate only accounts for changes in population size. It is assumed that the ASRs are normally distributed, and that the counts of cases or deaths in the crude rates are Poisson distributed. Thus, the resultant joinpoint trends of ASRs and crude rates may be different from each other and may include differences in the numbers of joinpoints in the optimal models, the locations of joinpoints, and the APCs. For example, in the analysis of the ASRs for male lung incidence, a joinpoint at 1987 was identified with the APC of the trend of 1987-2007 at −1.24%, while in the analysis of the crude rates the joinpoint was identified at 1990 with the APC of the trend in 1990-2007 at 0.24% (Figure 4). It should be noted that the population size grew from 1.8 million in 1975 to 3.2 million in 2007 and that the population aged, with an increase in the percentage of the population aged 65 and over from 7.4% in 1975 to 11.3% in 2007.

Figure 4. Optimal joinpoint trends for ASR and crude rates of Alberta male lung cancer incidence

(6) Crude rate vs counts:
When population size is constant over the study period, directly modeling counts is equivalent to modeling crude rates. In reality, population size and age structure are always varying. Thus, different results may be seen; different numbers of joinpoints, different locations of joinpoints, and/or different APCs of the trends. Again, for example, for male lung cancer incidence, modeling crude rates gave the optimal model with one joinpoint at 1990, but modeling counts gave one joinpoint at 1987. For cervical cancer incidence, the selected optimal model has no joinpoint for either the analysis of crude rates or
counts. Modeling crude rates showed the optimal model with declining APC = −0.5%, while modeling counts showed the optimal model with increasing APC = 1.15%. In general, directly modeling counts usually provides a larger value of APC due to population growth contribution.

(7) Zero counts:
For rare cancer outcomes, in particular cancer mortality of rare cancer sites, zero cases may occur in some years. In these instances, the Poisson model should be used to adjust the counts by the addition of a small constant (e.g. 0.5) to the zero counts. To see the impact of this addition, male thyroid cancer mortality data were analyzed; in 1995, there were no deaths. We modified the zero value by adding 0.1, 0.5 and 1, respectively. With population size as the offset, the results showed that there was no difference in the locations of joinpoints. There was no joinpoint for any of the optimal models for any of the three additions. The APCs are 1.84%, 1.83% and 1.83%, respectively.

(8) Age specific trend:
As noted above, Joinpoint software is only programmed to allow regression analysis with one independent variable. Thus, it is not possible to adjust for age in the overall model. Age specific trends, however, are available if the age variable is included in the “by variable’ tab.

Figure 5. Optimal joinpoint trends of age-specific rates in four age-groups for Alberta prostate cancer incidence. Blue: 70+, Red: 60-69, Green: 50-59, Black: 40-49

For example, prostate cancer was modeled using the four age-groups: 40-49, 50-59, 60-69, 70+ years. The age specific trends were tested by the pairwise difference in any two of four trends (Figure 5). Two
joinpoints were identified in the two youngest age groups (40-49), (50-59); one joinpoint was identified in the 60-69 year age group; and four joinpoints were identified in the oldest age group (70+ years). The trends in the two older age groups increased in the early years but decreased between 2002 and 2007, while the trends for the two youngest age groups increased over the whole time period. In comparison, only one joinpoint was identified in the analysis of the ASR, with a decreasing trend in the most recent segment. It is noted that age specific trend analysis may not be possible for rare cancers where there are a small number of cases/deaths in each age group. In these circumstances, data may be aggregated into larger age groups and/or calendar years.

(9) Different parameter setting:
In applying Joinpoint software, the different settings of parameters will lead to differences in the model selection. For example, when the default settings were used (minimum number of observations between two Joinpoints = 4, minimum number of observations from a joinpoint to either end of the data = 3, and maximum number of joinpoints = 4), four joinpoints were identified for the ASR of prostate cancer incidence. When the recommended settings were used (minimum number of observations between two Joinpoints = 6, minimum number of observations from a joinpoint to either end of the data = 6, and maximum number of joinpoints = 4), only one joinpoint was identified in the optimal model (see Fig. A5). The joinpoint analyzed results are sensitive to the parameter setting, although the joinpoint of the last segment was not changed in this example (see Figure 6).

Figure 6. Optimal joinpoint trends for ASRs of Alberta prostate cancer incidence using default setting (red lines) and suggested setting (blue lines)
Residual analysis for Joinpoint output:

Residual analysis can be used to compare the models for normally distributed data in relation to the choice of homoscedastic (unweighted) versus heteroscedastic (weighted) analysis, and the use of the logarithmic transformation. Although Pearson residuals are not provided by Joinpoint, they can be calculated by subtracting the observed data from the fitted data \( y_i - \hat{y}_i \) and the corresponding sample variance can be obtained by \( \hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \). The plot of normalized observations \( \frac{y_i}{\hat{\sigma}} \) against normalized fitted values \( \frac{\hat{y}_i}{\hat{\sigma}} \) can provide a visual check of the appropriateness of the transformation function,\(^{10}\) with confidence band as \( \frac{\hat{y}_i}{\hat{\sigma}} \pm 1.96 \). Using prostate cancer as an example, the normalized observations were plotted against normalized fitted values for the ASRs by four models (Figure 7): the original response scale (non-transformed) and constant variance (Panel a), original response scale and varying variance (inference by weighted least square) (Panel b), log-transformed response and constant variance (Panel c), and log-transformed response and varying variance (Panel d). Comparing Panel a with Panel c (under constant variance), and Panel b with Panel d (under varying variance), we see that there are two points out of the confidence band in Panel b, but one point out of the confidence band in Panel d. Therefore, it is necessary to take the log-transformation, particularly for the heterogeneous variance model. The Pearson residuals were plotted against the fitted values for the ASRs of prostate cancer incidence using the same four models (Figure 8). Distinctive “snake” patterns can be seen in Panel a and b, but those patterns were less distinct in Panel c and Panel d. Again, there are two points which were outside the confidence intervals \( (0 \pm 1.96) \) in Panel a, b and c, but only one point in Panel d. Given the results of these plots it can be concluded that the variances are heteroscedastic and the weighted model should be used with the logarithmic transformation for prostate cancer incidence.
Figure 7: Plots of normalized observations against normalized fitted values for ASRs of Alberta prostate cancer incidence
Figure 8: Plots of Pearson residuals against fitted values for ASRs of Alberta prostate cancer incidence
IV. Discussion and Recommendations

This trend analysis of the cancer incidence and mortality in the Alberta shows a favorable pattern over recent years in both sexes. The use of the joinpoint method has allowed a detailed and accurate description of the pattern of cancer incidence and mortality in recent years, since it identifies the calendar years in which statistically significant changes in trends occurred. This offers a clearer picture of actual trends in cancer incidence and mortality over long periods of time rather than using only one trend statistic.

It is not reasonable to expect that a single APC can accurately characterize the trend over an entire series of data. The joinpoint model uses statistical criteria to determine when and how often the APC changes. For cancer rates, it was fit using log-linear segments that are continuous at the joinpoints, so each segment can be characterized using an APC. For example, cancer rates may rise gradually for a period of several years, rise sharply for several years after that, then drop gradually for the next several years. A best fitted joinpoint model allows us to determine how long the APC remained constant, and when it changed.

Joinpoint software allows two kinds of probabilistic models: Normal and Poisson models. First, the software does not allow testing of the normality of the responses (e.g. ASRs), although this can be examined by the use of residual plots. If the response variable is not normal distributed, log-transformation may be a way to normalize the data and shrink the dispersion. Again, the normality of log-transformed data is still questionable. There are other transformation functions, e.g. the power transformation,\textsuperscript{11} which is more flexible for normalization. The Joinpoint software does not provide other transformations. Users should test the normalization of the response variable and search for the proper transformation before applying the software. For example, if log-transformation is not suitable, users may use another appropriate transformation and then enter the transformed data into the Joinpoint software. Be aware the transformation must be converted back properly when calculating EAPC. Secondly, weighted least square (WLS) inference method is a proper way to deal with heterogeneous variances in the linear model. The WLS also enhances the inference for the Poisson model, particularly when there is an over-dispersion problem in the count data. The Joinpoint statistical software does not include other more complicated statistical models, for example, Poisson mixture and negative binomial models, to capture the over-dispersion problem. Finally, it is noted that the serial correlation among the time series data is not taken into consideration for statistical inference in the current version of Joinpoint software. Cancer incidence and mortality rates may not be rigorous time series data as the inherent population changes over time, e.g. births, deaths, and migration. The data may be considered to be a time series, however, if the
population is regarded as a whole cohort. Generally, the observed data are not considered to be independent of each other. The Joinpoint software works for discrete time only and does not treat the data as a time series.

Small numbers of cases were observed for small populations and/or rare cancer sites. In general, where there are small values (number of cases or ASR), it is difficult to detect any statistically significant joinpoint in the trends due to the large dispersion of the data distribution. In a situation where no event occurred in a certain year, to avoid zero counts for log-link function in Poisson model, 0.5 is added to all zero value in counts. It should be noted that for rare cancers, this adjustment would potentially cause a bias in the estimates, as 0.5 is relatively large compared to counts less than 5. One may modify the data base such that the scale of the numbers is larger, for example, multiplying the number of cases and/or population by 10; 0.5 would be relatively small compared with 10 or 20. When analyzing rates, zero rates would make it impossible to log-transform the rates, and would be problematic when fitting the model by weighted least square method as the relevant standard errors of ASRs are not applicable. One may use original ASRs (without log-transformation) and assume the constant variance in the linear model. Alternatively, one can replace the zero ASR and its weight by very small values, say 0.000001, before the data are input into Joinpoint software.

Joinpoint software provides two methods for model selection: Permutation Test and Bayesian Information Criteria (BIC). The permutation test uses a series of Monte Carlo simulation procedures, which may be time consuming but gives a conservative testing result; while the BIC gives an aggressive testing result with a very fast computation, as it is simple to calculate the value of BIC (log likelihood value – the penalty of extra parameters). BIC method usually gives the same optimal model as permutation test method, or gives more segments with significant APC.

Parameter setting needs skill, as was seen in the Results section. Different settings may lead to different joinpoint trends. For the purpose of projections, we expect that joinpoint trend analysis would provide meaningful and substantial trends for the last segment. Users may set different, but reasonable, (i) maximum number of joinpoints, (ii) minimum number of observations between two joinpoints, and (iii) minimum number of observations from a joinpoint to either end of the data for different cancer sites according to their purposes and experience.
From joinpoint analysis, the most recent trend can be identified and used to predict future cancer incidence and mortality rates. For example, the trend of prostate cancer incidence in 2001 – 2007 can be used to project future rates. The APC is -3.96%, and the rate is 125.72 per 100,000 in 2007, the rate is 125.72 x (1-0.396) = 125.72 × 0.9604 = 120.74 in 2008 and 120.74 × 0.9604 = 115.96 in 2009.

For some cancer sites, the trend in the last segment for the optimal model may not be statistically significant. In these cases, either the average rates for this period may be used to predict future rates, or the estimated trend may be used. If the non significant estimated trend is used, this may lead to wide confidence intervals for the projections. Alternatively, it may be preferable to use a Joinpoint model to predict future rates if there is other evidence of changing rates over time. Again the confidence interval may be wide.

Joinpoint is often used for short term projections by extrapolating the trend in the age standardized rates and applying the total annual population for a given year to calculate the projected number of cases or deaths. Short-term cancer projections of the number of cases or deaths may be impacted, however, by changes in population size and structure e.g. aging. The projections based on the last segment from joinpoint analysis for age-standardized rates may be seriously under-estimated if the population in the projection year is older than that for the standard population, (usually the 1991 Canadian population), as aging effects can not be taken into consideration in joinpoint calendar yearly trend analysis simultaneously. Joinpoint software is a statistical package for the simple regression model and works for one data series at a time. One way to overcome this issue is to standardize the cancer rates to the population structure of the projection year; however, this is burdensome if projections are required for several years. Also, the results from these analyses may indicate different joinpoints, and different trends among the analyses for the different projection years. An alternate solution would be to use the Poisson simple regression model for the yearly number of cases with population as the offset when only the projection of numbers is of interest. In conclusion, Joinpoint is a powerful tool to investigate changes in trends of cancer incidence and mortality.
References:

1. National Cancer Institute. Joinpoint Regression Program. version 3.4.2, 2009,
   (Available at http://srab.cancer.gov/joinpoint/)
3. Qiu D, Katanoda K, Marugame T and Sobue T. A Joinpoint regression analysis of long-term trends in
   (Available at http://progressreport.cancer.gov/)
   2006, featuring colorectal cancer trends and impact of interventions (risk factors, screening, and
   Published online: www.interscience.wiley.com
8. Hudson DJ. Fitting segmented curves whose join points have to be estimated. Journal of American
9. Kim HJ, Fay MP, Feuer EJ and Midthune DN. Permutation tests for joinpoint regression with

Fig. A1. All Cancer: Observed Incidence/Mortality Rates and Estimated Trends from Joinpoint Analysis; Blue: Male, Red: Female; Left Panel: Incidence, Right Panel: Mortality. (Same below)

Fig. A2. Colorectal Cancer

Fig. A3. Lung Cancer
Fig. A4. Female Breast Cancer

Fig. A5. Prostate Cancer

Fig. A6. Oral Cancer
Fig. A7. Bladder Cancer

Fig. A8. Brain Cancer

Fig. A9. Cervix Cancer
Fig. A10. Esophageal Cancer

Fig. A11. Hodgkin Lymphoma

Fig. A12. Kidney Cancer
Fig. A13. Larynx Cancer

Fig. A14. Leukemia

Fig. A15. Liver Cancer
Fig. A16. Melanoma of Skin

Fig. A17. Multiple Myeloma

Fig. A18. Non-Hodgkin’s Lymphoma
Fig. A19. Ovary Cancer

Fig. A20. Pancreas Cancer

Fig. A21. Stomach Cancer
Fig. A22. Testis Cancer

Fig. A23. Thyroid Cancer

Fig. A24. Body of Uterus Cancer